ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Recognition of Electronic Disguised Voices by the Means of MFCC

Sachin Kurian[1], Nikhil.G.Kurup[2]

PG Student [SP], Dept. of ECE, College of Engineering, Kallooppara, Kerala, India[1]

Assistant Professor, Dept. of ECE, College of Engineering, Kallooppara, Kerala, India[2]

ABSTRACT: Voice disguise is a deliberate action of a speaker who wants to falsify or to conceal his/her identity. Since voice disguise has great negative impact on establishing authenticity of audio evidence in forensics, and has shown an increasing tendency in illegal applications, it is important to identify whether a suspected voice has been disguised or not. However, few studies on such identification have been reported. In this paper, we propose an algorithm to recognize electronic disguised voices by the means of MFCC. Statistical moments of Mel-frequency cepstrum coefficients (MFCC) including mean values and correlation coefficients are extracted as acoustic features. Then the pitch and formant calculation is done. After that an algorithm based on support vector machine is used to separate original voice from the disguised voices is proposed. Then an approach for detection of disguised voice based on the extracted features and Hidden Markov Model (HMM) classifiers is also proposed.

KEYWORDS: Mel Frequency Cepstral Coefficients(MFCC), Feature Extraction, Support Vector Machine(SVM), Hidden Markov Model(HMM).

I.INTRODUCTION

Voice disguise refers to a system of altering a person's voice to either make them sound like someone else or to disguise their voice. Voice disguise methods can be divided into two types: non electronic disguise and electronic disguise. Non-electronic disguise is used to alter the voice tone of a speaker by disturbing his human speech production system mechanically. Common non-electronic methods include pinching the nostrils, clenching the jaw, using a bite block, pulling the cheek, holding the tongue, speaking with an object in mouth, etc.An electronic disguised voice is obtained by using electronic scrambling devices to modify frequency spectral properties such as the voice pitch and voice formants of an original voice. As a result, criminal cases using electronic disguise have been increasing in phone communications, online chatting, and other speech applications in recent years. Hence, detection of electronic disguised voice has become an important and emergent issue.

Voice disguise is an intentional operation to conceal or forge speaker's identity by changing his or her voice tone.The principle of voice disguise is to raise or to lower voice pitch by stretching or compressing frequency spectrum. Nowadays, an increasing number of audio editing softwares, such as Audacity [3], Cool Edit [4], and PRAAT [5] provide disguising tools with a variety of disguising factors. the applications based on electronic disguise are widely used in privacy protection, entertainment, speech synthesis, speech coding, and other fields. We focus on electronic disguised voices instead of non-electronic methods

Mel-Frequency Cepstrum Coefficients (MFCC) are widely used as acoustic features in speaker recognition. Voice disguise modifies the frequency spectral properties of an original voice and thus changes the MFCC of the voice. Hence, MFCC statistical moments can be used for the identification of disguised voices. MFCC statistical moments including mean values and correlation coefficients are extracted as acoustic features. Then, an algorithm based on the extracted features and hidden markov models classifiers is proposed to separate disguised voices from original voices. An identification system based on HMM classifiers is designed in our work. The basic idea of the proposed algorithm is that it is possible to distinguish the voices disguised by a certain factor from original voices.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

II.LITERATURE SURVEY

Research on voice disguise started in the 1970s with phoneticians like Künzel, Koester, and it is really over the past 10 years that researchers have tried to develop automatic systems to detect the disguise. This question of voice disguise in forensic sciences was not very developed in the literature, certainly because of the difficulties to distinguish a normal voice from a disguised voice in criminal applications. Nevertheless, the increase of voice use in multimedia applications and the current performance of speaker recognition systems offer a new interest for voice disguise**.**Natural and disguised speech data from 100 German speakers recorded 5 times over a period of 7 to 9 months were used in a series of speaker recognition experiments. Results indicate that the three types of voice disguise selected affect the performance of the system only marginally if reference populations contain speech data which exhibit the same type of disguise.[2]

In order to evaluate the best way to detect automatically disguised voice, three different classifications have been used on the previously described features:k-nearest-neighbors,GMM (Gaussian Mixture Model),VO (Vector quantization) and SVM (Support vector machine)[3].Voice disguise can be classified into two categories: voice conversion and voice transformation. Voice conversion is to transform one's voice to imitate a target person provided with the target's acoustic information, while voice transformation is to change the sound without any target. Voice transformation can be implemented by non-electronic and electronic means. Non-electronic means includes the alteration of voice by using a mechanic system like a mask over the mouth, a pen in the mouth or pinching the nostril.[4]

we propose an algorithm to identify electronic disguised voices. Since voice disguise, in essence, the modification of the frequency spectrum of speech signals, and mel-frequency cepstrum coefficients (MFCCs) can be used to well describe frequency spectral properties, MFCC-based features are supposed to be effective for the identification of disguised voices. In this paper, MFCC statistical moments including mean values and correlation coefficients are extracted as acoustic features. Then, an algorithm based on the extracted features and support vector machine classifiers is proposed to separate disguised voices from original voices.[1]

III.PRPPOSED METHOD

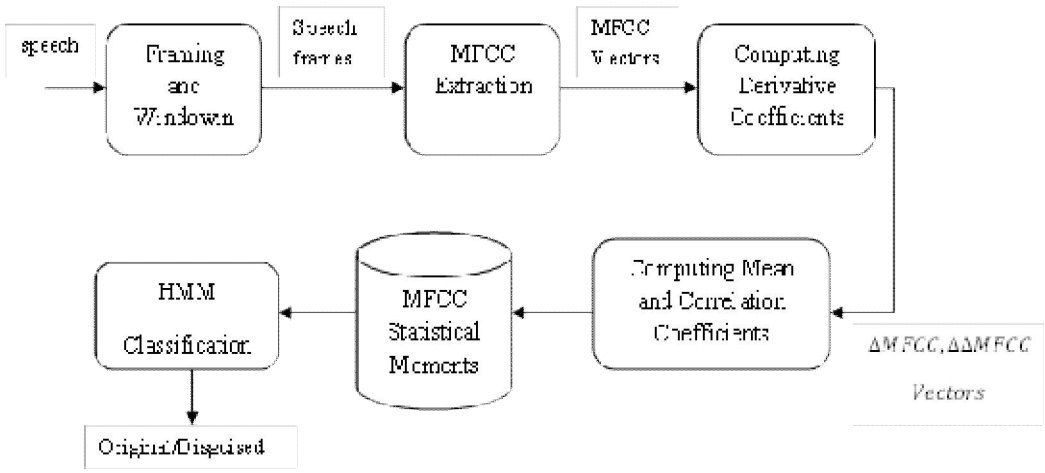

Fig 1: Block diagram

Since MFCC can be used to model the human auditory perception performed in the inner ear, they are widely used to represent speech signals in speech recognition, speaker recognition, and other speech applications. Traditional MFCC are extracted as follows:

a) PRE-EMPHASIS

Spectrum for voiced segments has more energy at lower frequencies than higher frequencies .This process is called spectral tilt. pre-emphasis is the process of boosting the energy in the high frequencies.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

b) FRAME BLOCKING

An audio signal is constantly changing, so to simplify things we assume that on short time scales the audio signal doesn't change much . This is why we frame the signal into 20-40ms frames.

c) WINDOWING

It minimizes the spectral distortion.Windowed signal is considered as stationary.Typically hamming window is used.

c) FFT

. It transforms each frame of N samples from time domain to frequency domain. FFT is the fast algorithm to implement DFT. Resulting signal after DFT is spectrum.

d) MEL SCALE

Human ear does not follow linear frequency spacing. Frequency in Hz converted in to a scale called mel-sacle. A mel is a unit of pitch. Mel scale is approximately linear below 1KHz and logarithmic above 1 KHz. The Mel scale tells us exactly how to space our filterbanks and how wide to make them.[3]

$$Mel(f) = 2595\log_{10}(1+\frac{f}{700}) \quad (1)$$

e) MEL FILTER BANK

The human auditory system doesn't interpret pitch in a linear manner. It is uniformly spaced before 1KHz and logarithmic above 1KHz. Power spectrum is filtered mel-filter bank. Mel- filter bank consist of triangular band-pass filters.

DERIVATIVE COEFFICIENTS

Derivative coefficients (ΔMFCC and ΔΔMFCC) reflecting dynamic cepstral features are computed from the MFCC vectors. As a result, *N L*-dimensional ΔMFCC vectors and *N L*-dimensional ΔΔMFCC vectors are obtained. Also known as differential and acceleration coefficients.

STATISTICAL MOMENTS

Two kinds of statistical moments, including the mean values *E j* of each component set *Vj*, and the correlation coefficients *CRj j* ' between different component sets *Vj*and *Vj*'are taken into consideration. They are calculated by :[5]

$$E\,j = E(Vj\,),\ j = 1,2,.....L \quad (2)$$

$$CR_{j,j'} = \frac{cov(V_j\ , V_j')}{\sqrt{VAR(V_j)}\sqrt{VAR(V_j')}} \quad (3)$$

$$W = [W_{MFCC}, W_{\Delta MFCC}, W_{\Delta\Delta MFCC}] \quad (4)$$

HMM CLASSIFICATION

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be presented as the simplest dynamic Bayesian network. The mathematics behind the HMM were developed by L. E. Baum and coworkers. It is closely related to an earlier work on the optimal nonlinear filtering problem by Ruslan L. Stratonovich. who was the first to describe the forward-backward procedure.

Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharge and bioinformatics.

A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

process rather than independent of each other. Recently, hidden Markov models have been generalized to pairwise Markov models and triplet Markov models which allow consideration of more complex data structures and the modelling of non-stationary data.[5]

PITCH AND FORMANT CALCULATION

Pitch and formant frequencies are important features in speech. The analog signal is converted to digital by sampling with a suitable rate and quantized. The digital signal is thenhamming windowed to convert it into a suitable frame size. The signal is converted into frequency domain by using Fast Fourier Transform. The absolute values of the signal are considered and then the logarithm of the signal is obtained. The signal is then transformed into Cepstral domain by taking its IFFT. The very first signal peak represents the pitch frequency. Formants are defined as the spectral peaks of sound spectrum, of the voice, of a person. In speech science and phonetics, formant frequencies refer to the acoustic resonance of the human vocal tract [6]. They are often measured as amplitude peaks in the frequency spectrum of the sound wave. The Linear predictive coding technique (LPC) has been used for estimation of the formant frequencies [3]. The analog signal is converted in .wav digital format. The signal is transformed to frequency domain using FFT and the power spectrum is further calculated. Then the signal is passed through a Linear Predictive Filter (LPC).

IV.RESULT AND DISCUSSION

- The data base for the project is taken from UME_ERJ.
- First the original and disguised signals are undergone training.
- During the training phase, the power spectrum, cepstrum, derivatives of original and disguised voices are calculated.

The cepstrum of original signal and the disgused signal is as shown in Fig 2&Fig 3:

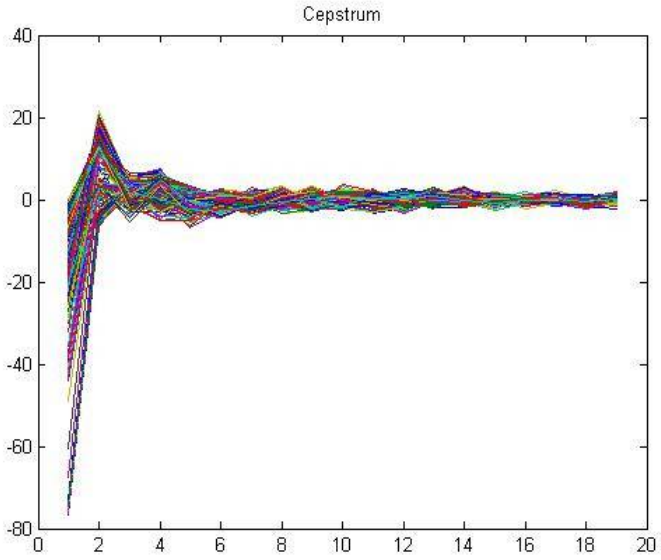

Fig 2 Cepstrum of original voice

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

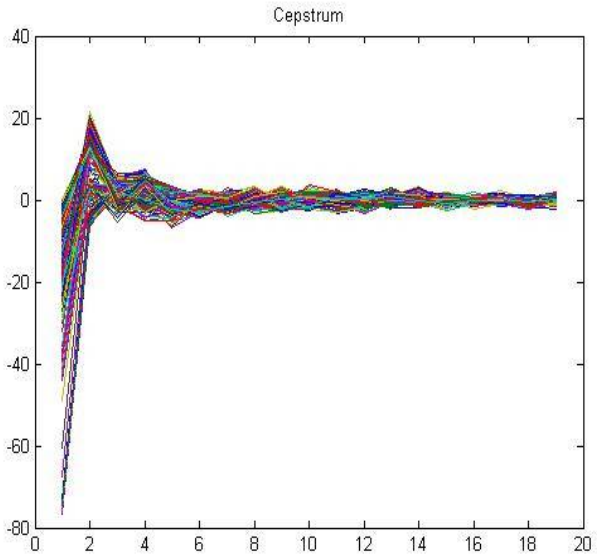

Fig 3Cepstrum of disguised voice

VOICE DETECTION ACCURACY RATE

Voice detection accuracy of the identification of original and disguised voices is evaluated using the ratio of correctly recognized voice against the total number of voice tested. Voice accuracy of disguised and original voice are tested. The Table 1 shows the voice detection accuracy

TESTING VOICES	NO. OF VOICES TESTED	CORRECTLY CLASSIFIED
DISGUISED VOICES	20	20
ORIGINAL VOICES	20	16

Table1: Verified Output using SVM (Without Modified Features)

TESTING VOICES	NO. OF VOICES TESTED	CORRECTLY CLASSIFIED
DISGUISED VOICES	20	20
ORIGINAL VOICES	20	18

Table 2 : Verified Output using SVM (With Modified Features)

TESTING VOICES	NO. OF VOICES TESTED	CORRECTLY CLASSIFIED
DISGUISED VOICES	20	20
ORIGINAL VOICES	20	19

Table 3 : Verified Output using HMM

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Overall Accuracy

The overall voice accuracy is calculated as the ratio of the total number of voice recognized to the total number of digits tested.The overall voice accuracy is shown below :

Using SVM without modified features	92.5%
Using SVM with modified features	95%
Uing HMM	97.5%

Table 4 : Overall accuracy rate table

VI.CONCLUSION

Here, an algorithm for identifying electronic disguised voices is proposed. MFCC statistical moments, i.e., the mean values and correlation coefficients of MFCC vectors, derivatives, are extracted as acoustic features. And also pitch and formant features are also calculated. Here SVM classifier and used to separate original voice from disguised voices. A statistical analysis of the acoustic features indicates that the distributions of the feature components of original voices are altered due to voice disguise. Thus these acoustic features can be used to separate disguised voices from original voices. An identification system based on HMM classifiers is designed in our work. The basic idea of the proposed algorithm is that it is possible to distinguish the voices disguised by a certain factor from original voices. From a practical point of view, we think that the proposed algorithm of identifying disguised voices can help in forensics and security to some extent, such as in narrowing the scope of suspects or in providing an early warning about a suspected v

REFERENCES

[1] Haojun Wu, Student Member, IEEE, Yong Wang, Member, IEEE, and Jiwu Huang, Senior Member, IEEE " Identification of Electronic Disguised Voices'' IEEE Transactions On Information Forensics And Security, Vol. 9, No. 3, March 2014.
[2] H. J. Künzel, J. Gonzalez-Rodriguez, and J. Ortega-García, "Effect of voice disguise on the performance of a forensic automatic speaker recognition system," in Proc. IEEE Int. Workshop Speaker Lang. Recognit., Jun. 2004, pp. 1–4.
[3] P. Perrot and G. Chollet, "The question of disguised voice," J. Acoust. Soc. Amer., vol. 123, no. 5, pp. 3878-1–3878-5, Jun. 2008
[4] Y. Wang, Y. Deng, H. Wu, and J. Huang, "Blind detection of electronic voice transformation with natural disguise," in Proc. Int. Workshop Digital Forensics Watermarking, 2012, LNCS 7809, pp. 336–343.
[5] H. Wu, Y. Wang, and J. Huang, "Blind detection of electronic disguised voice," in Proc. IEEE ICASSP, vol. 1. Feb. 2013, pp. 3013–3017.
[6] C. Zhang and T. Tan, "Voice disguise and automatic speaker recognition," Forensic Sci. Int., vol. 175, no. 2, pp. 118–122, 2008.
[7] T. Tan, "The effect of voice disguise on automatic speaker recognition," in Proc. IEEE Int. CISP, vol. 8. Oct. 2010, pp. 3538–3541.
[8] S. S. Kajarekar, H. Bratt, E. Shriberg, and R. de Leon, "A study of intentional voice modifications for evading automatic speaker recognition,"in Proc. IEEE Int. Workshop Speaker Lang. Recognit., Jun. 2006, pp. 1–6.
[9] S. Roucos and A. Wilgus, "High quality time-scale modification for speech," in Proc. IEEE ICASSP, vol. 10. Apr. 1985, pp. 493–496.
[10]J. Laroche, "Time and pitch scale modification of audio signals," in Applications of Digital Signal Processing to Audio and Acoustics. New York, NY, USA: Springer-Verlag, 2002, pp. 279–309.
[11]R. E. Crochiere and L. R. Rabiner, "Interpolation and decimation of digital signals—A tutorial review," Proc. IEEE, vol. 69, no. 3 pp. 300–331, Mar. 1981.
[12]S. E. Trehub, A. J. Cohen, L. A. Thorpe, and B. A. Morrongiello, "Development of the perception of musical relations: Semitone and diatonic structure," Development, vol. 12, no. 3, pp. 295–301, 1986.
[13] X. Zhu, G. Beauregard, and L. Wyse, "Real-time signal estimation from modified short-time fourier transform magnitude spectra," IEEE Trans. Audio, Speech Lang. Process., vol. 15, no. 5, pp. 1645–1653, Aug. 2007.
[14] J. Gonzalez-Rodriguez, D. Ramos-Castro, M. Garcia-Gomar, and J. Ortega-Garcia, "On robust estimation of likelihood ratios: TheATVS-UPM system at 2003 NFI/TNO forensic evaluation," inProc. IEEE Int. Workshop Speaker Language Recognit., Jun. 2004,pp. 1–8.

BIOGRAPHY

Sachin Kurian is currently pursuing M.Tech in electronics with specialisation in SignalProcessing from College Of Engineering, Kallooppara (CUSAT University), Kerala, India. He received her B.Tech degree in Electronics and Communication from Ponjesly College of Engineering and Technology, Nagercoil (Anna University), Tamil Nadu, India. His areas of interest are Digital Image Processing, Speech Signal Processing, Digital Communication, Wavelet and Embedded Design.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Nikhil.G.Kurup is currently working as Assistant Professor in College Of Engineering,Kallooppara, Kerala, India. He received her M.Tech on Communication Engineering from College of Engineering, Cherthala and B.Tech degree from College of Engineering Kanunagapally, Kerala, India. His areas of interest are Digital Signal Processing, Biometrics, Digital Communication, Wavelet, Wireless Technology and Embedded Design.